Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference

C. P. Van Tassell and L. D. Van Vleck

The online version of this article, along with updated information and services, is located on the World Wide Web at:

http://jas.fass.org

Multiple-Trait Gibbs Sampler for Animal Models: Flexible Programs for Bayesian and Likelihood-Based (Co)Variance Component Inference

C. P. Van Tassell[1] and L. D. Van Vleck

Roman L. Hruska U.S. Meat Animal Research Center, USDA-ARS,
University of Nebraska, Lincoln 68583-0908

ABSTRACT: A set of FORTRAN programs to implement a multiple-trait Gibbs sampling algorithm for (co)variance component inference in animal models (MTGSAM) was developed. The MTGSAM programs are available to the public. The programs support models with correlated genetic effects and arbitrary numbers of covariates, fixed effects, and independent random effects for each trait. Any combination of missing traits is allowed. The programs were used to estimate variance components for 50 replicates of simulated data. Each replicate consisted of 50 animals of each sex in each of four generations, for 400 animals in each replicate for two traits. For MTGSAM, informative prior distributions for variance components were inverted Wishart random variables with 10 df and means equal to the simulation parameters. A total of 15,000 Gibbs sampling rounds were completed for each replicate, with 2,000 rounds discarded for burn-in. For multiple-trait derivative free restricted maximum likelihood (MTDFREML), starting values for the variance components were the simulation parameters. Averages of posterior mean of variance components estimated using MTGSAM with informative and flat prior distributions for variance components and REML estimates obtained using MTDFREML indicated that all three methods were empirically unbiased. Correlations between estimates from MTGSAM using flat priors and MTDFREML all exceeded .99.

Key Words: Gibbs Sampler, Variance Components, Bayesian Theory, Computer Software, Genetics

Introduction

Estimation of variance components (**VC**) has long been important in animal breeding. Accurate estimates of VC are important because prediction error variances for predicted random effects (e.g., breeding values) increase as differences between estimated and true values of VC increase (Henderson, 1975; Schaeffer, 1984). Currently, REML is considered the method of choice for estimating VC (Meyer, 1990). The use of REML in animal breeding has increased dramatically due to availability of general purpose programs such as DFREML (Meyer, 1988), MTDFREML (Boldman et al., 1995), and VCE (Groeneveld, 1994). The multiple-trait Gibbs sampler for animal models (**MTGSAM**) programs have been developed to implement the Gibbs sampling (**GS**) algorithm for Bayesian analysis of a broad range of animal models (Van Tassell and Van Vleck, 1995). These programs expand the methods available for statistical analysis of animal breeding data.

The objectives of this study were to derive and present the theoretical results necessary to implement the Gibbs sampler for a general multiple-trait animal model and to compare posterior mean estimates obtained using different prior distributions with those obtained using REML for simulated data.

Methods and Materials

Gibbs Sampling and Bayesian Concepts

Gibbs sampling is a Monte Carlo numerical integration method that allows inferences to be made about joint or marginal distributions, even if appropriate densities cannot be explicitly formed (Geman and Geman, 1984). The use of GS allows analysis of Bayesian posterior distributions that had been computationally intractable due to the numerical integration required to obtain those distributions. The GS algorithm is applied by generating a realized value for

---

[1]To whom correspondence should be addressed. Current address: USDA-ARS-AIPL, Building 263, BARC-East, 10300 Baltimore Ave., Beltsville, MD 20705-2350. Phone: (301)504-9271; FAX: (301)504-8092; E-mail: curtvt@aipl.arsusda.gov.

each unknown parameter (considered a random variable in a Bayesian framework) in a model, in turn, from the fully conditional distribution of that variable with all other parameters in the model and the data considered known. The cycle of generating each parameter is repeated. Eventually, the Gibbs sampler converges to the posterior distribution, and the values drawn after that convergence are considered random samples from the posterior distribution. The number of rounds discarded before the values are considered samples from the posterior distribution is usually called the burn-in period. There are typically large positive correlations between consecutive samples for a parameter drawn from the Gibbs chain. Independent samples of the parameters are usually needed for posterior density estimation and can also be used for estimation of Monte Carlo error. The number of rounds between samples needed to obtain uncorrelated samples is called the thinning interval. The Monte Carlo error can also be estimated using correlated samples. Sorensen et al. (1995) describe calculation of Monte Carlo error. The Monte Carlo error is the error in parameter estimation due to the number of samples used from the Gibbs chain. The Monte Carlo error is inversely proportional to the length of the Gibbs chain.

Gibbs sampling has been used for Bayesian inference in animal breeding problems in several applications, including VC estimation in sire models (Wang et al., 1993), animal models (Wang et al., 1994b; Van Tassell et al., 1995), and maternal effects models (Jensen et al., 1994; Van Tassell, 1994). Gibbs sampling has also been applied to Bayesian inference about response to selection (Sorensen et al., 1994; Wang et al., 1994a). Recently, GS has been extended to Bayesian inference in threshold models for categorical data (Albert and Chib, 1993; Sorensen et al., 1995). Each of these cases involved development of computer programs needed for the particular model.

It is important to note that likelihood inference can be viewed as a special case of the more general Bayesian inference. With normality, the posterior distribution is simply the (frequentist) likelihood function scaled by prior distributions of the unknown parameters in the model. When a flat prior is used for the parameters, then the posterior distribution is the likelihood of the parameters given the data. A flat prior distribution is one that attempts to represent a lack of prior knowledge; in that case the prior distribution is proportional to a constant. Gibbs sampling can be applied characterizing both the posterior distribution in a Bayesian inference and the likelihood function in a frequentist inference (Tanner, 1993). Estimates of posterior means of VC with flat prior distributions for fixed effects and variance components correspond to the VC estimates from integrated likelihood (**VEIL**) as described by Gianola and Foulley (1990).

Program Flexibility

The MTGSAM programs support a variety of models. The model for each trait can include any number of fixed and uncorrelated random effects. In addition to the additive genetic effect (i.e., animal effect), an additional correlated genetic effect (e.g., maternal genetic effect) can be included for each trait. The programs support multiple trait models with missing observations. The possible models correspond to those allowed by the multiple-trait derivative-free REML (**MTDFREML**) programs. Like the MTDFREML programs, the MTGSAM programs can be used for sire models by considering the genetic effect as a sire effect, and a sire-maternal grandsire model can be fit by considering maternal grandsire effect as the correlated genetic effect. The MTGSAM programs generate the Gibbs samples for those models and calculate estimates of posterior means of the associated parameters.

Model Assumptions

In matrix notation, the mixed linear model equation for the N × 1 observation vector, **y**, is:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{e},$$

where β is the vector of fixed effects associated with records in **y** by **X**, and **u** is the vector of random effects associated with records in **y** by **Z**, and **e** is the vector of random residual effects. Completing the model definition, assume

$$\mathrm{E}\begin{bmatrix}\mathbf{y}\\\mathbf{u}\\\mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{X}\beta\\\mathbf{0}\\\mathbf{0}\end{bmatrix} \text{ and}$$

$$\mathrm{Var}\begin{bmatrix}\mathbf{y}\\\mathbf{u}\\\mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{Z}\Sigma\mathbf{Z}' + \mathbf{R} & \mathbf{Z}\Sigma & \mathbf{R}\\\Sigma\mathbf{Z}' & \Sigma & \mathbf{0}\\\mathbf{R} & \mathbf{0} & \mathbf{R}\end{bmatrix}.$$

There is no way to generate covariance matrices from the fully conditional distribution of the VC with specific covariances restricted to zero. Therefore, uncorrelated random and residual effects are grouped by traits to impose zero covariances among effects in order to implement the GS algorithm. The covariances between effects in different groups are assumed to be zero. The subdivision of uncorrelated random and residual effects into groups usually corresponds to groups of traits that are not observed on the same animal, e.g., the effects on sex-limited traits such as milk production and scrotal circumference. If zero covariances are desired between two factors, then the factors need to be considered as members of different groups. The use of the term group refers only to the concept of assigning random effects to different categories for the sake of VC estimation and is unrelated to the idea of genetic grouping.

Let $\mathbf{u}' = \begin{bmatrix} \mathbf{u}'_0 & \mathbf{u}'_1 & ... & \mathbf{u}'_\gamma \end{bmatrix}$, where $\mathbf{u}_0$ corresponds to the additive genetic effects, $\mathbf{u}_i$ corresponds to the uncorrelated random effects in group i for $i > 0$, and γ is the number of groups of uncorrelated random effects. The vector $\mathbf{u}_0$ is sorted by animal within trait and the traits ordered with direct genetic effects in order for each trait followed by the additional correlated genetic effects for traits that have them. Then $\mathbf{G}$ is a $d_0 \times d_0$ matrix that describes the genetic (co)variances among traits for an animal, where d_0 is the number of traits plus the number of traits with correlated genetic effects. Let $\mathbf{A}$ be an $n \times n$ matrix describing the covariance structure among genetic effects within traits; typically $\mathbf{A}$ is the numerator relationship matrix, and n is the number of animals. The vector $\mathbf{u}_i$ corresponds to random effects that are uncorrelated with the genetic effects and the other random effects (e.g., permanent environmental effects). The effects are sorted by level within the traits that include that effect. Only random effects with the same levels (i.e., coded by the same values) across traits can be included in a group of effects with non-zero covariances among those effects. Then $\mathbf{D}_i$ is the $d_i \times d_i$ matrix of (co)variances among the random effects across traits in group i for an animal, where d_i is the number of uncorrelated random effects in group i, i.e., the number of traits represented in group i. Let n_i be the number of levels in each of the d_i uncorrelated random effects in group i. Based on the partitioned form and the assumptions, the matrix Σ can be written as

$$\Sigma = \begin{bmatrix} (\mathbf{G} \otimes \mathbf{A}) & \mathbf{0} \\ \mathbf{0} & \overset{\gamma}{\underset{i=1}{\oplus}} \begin{bmatrix} \mathbf{D}_i \otimes \mathbf{I}_{n_i} \end{bmatrix} \end{bmatrix},$$

where $\otimes$ and $\oplus$ correspond to direct product and direct sum operators, respectively (see Searle, 1982, for a description of these operators). Let $\Sigma_0 = \mathbf{G} \otimes \mathbf{A}$ and $\Sigma_i = \mathbf{D}_i \otimes \mathbf{I}_{n_i}$; then Σ can be also written as $\Sigma = \overset{\gamma}{\underset{i=0}{\oplus}}\Sigma_i$.

An animal may have multiple observations. The residual effects are assumed to be uncorrelated for the different records across the same animal, although residual effects for traits within a record may be correlated. If the data vector, $\mathbf{y}$, is assumed to be sorted by trait within record within animal, then $\mathbf{R}$ is a block diagonal matrix

$$\mathbf{R} = \overset{q}{\underset{i=1}{\oplus}} \mathbf{R}^*_j,$$

where $\mathbf{R}^*_j$ is an $r_j \times r_j$ matrix of (co)variances of residuals for the traits measured on record j, q is the

number of records, and r_j is the number of traits observed for record j. Note that for an animal with all traits measured $\mathbf{R}^*_j = \overset{\rho}{\underset{i=1}{\oplus}} \mathbf{R}_i$, where $\mathbf{R}_i$ is the $t_i \times t_i$ covariance matrix among residuals for group i of the residual effects, t_i is the number of traits in residual group i, and ρ is the number of groups of residual effects.

Henderson's mixed model equations (**MME**) (1984) simplify for many situations the calculation of best linear unbiased estimators of β, $\hat{\beta}$, and best linear unbiased predictors of $\mathbf{u}$, $\hat{\mathbf{u}}$. In general form the MME are

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \Sigma^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}.$$

The equations will also be written as $\mathbf{C}\hat{\mathbf{s}} = \mathbf{r}$. Although $\mathbf{R}$ is of order equal to the number of records, $\mathbf{R}$ is often assumed to be diagonal for single-trait analyses, often $\mathbf{I}\sigma_e^2$, and block diagonal (blocks of order equal to the number of traits) for multiple-trait analyses, so that calculations involving $\mathbf{R}^{-1}$ are easy.

The GS algorithm requires that the joint distribution of the parameters conditional on the data is proper, i.e., that the distribution integrates to a finite value. In some models and data structures the integral of the likelihood is infinite, and application of the Gibbs sampler to a model with flat priors may yield incorrect results (Hobert, 1994). Even when the integral of the likelihood is infinite, however, use of an appropriate prior distribution will result in a model with a proper posterior distribution. The posterior distribution will always be proper when proper prior distributions are used for all parameters in the model (including fixed effects), and there are many cases where flat or other improper priors can be used and still result in a proper posterior distribution. Hobert (1994) investigated this problem for the mixed linear model with one random effect. The problem has not been well characterized for more complex models, especially multiple trait models. In cases where flat prior distributions can safely be used for the VC, the posterior mean estimates correspond to VEIL (Gianola and Foulley, 1990) estimates for the parameters and the posterior distributions of parameters correspond to marginal likelihoods (Tanner, 1993). For generality, the notation and derivations will be for the Bayesian model.

Prior Distributions. To specify fully the Bayesian model with unknown VC, additional assumptions must be made. The MTGSAM programs were developed assuming flat prior distributions for the fixed effects. The random effects are assumed to be normally distributed. For the genetic effects there will be an additional assumption of a known covariance structure corresponding to the numerator relationship

matrix. Finally, the residual effects are assumed to be distributed normally. These assumptions are those used with most likelihood-based methods when these methods are interpreted from the Bayesian point of view. In addition, when VC are known, the posterior distribution under these assumptions is multivariate normal with the mean equal to $\hat{\beta}$ and $\hat{\mathbf{u}}$ and variance-covariance matrix $\mathbf{C}^{-1}$ (Gianola and Fernando, 1986; Gianola et al., 1990).

The animal and correlated genetic effects are assumed to have non-zero covariances that can be estimated for all trait combinations. Non-zero covariances are allowed only among the same random effects for different traits. In addition, the user of MTGSAM can specify that some of these covariances are to be restricted to zero by specifying additional groups of effects.

Inverted Wishart (**IW**) distributions are used as prior distributions for the (co)variance components mainly for computational simplicity. The IW prior distributions for the matrix of (co)variance components are assumed to be mutually independent. The Wishart density describes the distribution of sums of squares and cross-products of standard normal random variables (**RV**) (Odell and Feiveson, 1966), and if **B** is a Wishart RV, then $\mathbf{T} = \mathbf{B}^{-1}$ is an IW RV. In the univariate case, this corresponds to an inverted chi-square distribution. Although other prior distributions could be used, the GS algorithm would be more complex.

If **T** is distributed as an IW variable, in other words $\mathbf{T} \sim \text{IW}(\mathbf{T}_0, \nu)$, then the form of the IW distribution is

$$f(\mathbf{T} \mid \mathbf{T}_0, \nu) \;\propto\; |\mathbf{T}|^{-\frac{1}{2}(\nu+m+1)} \times \exp\left(\text{tr}\left(-\frac{1}{2}\mathbf{T}_0^{-1}\mathbf{T}^{-1}\right)\right),$$

for **T**, $\mathbf{T}_0$ positive definite and $\nu > m + 1$, where $\mathbf{T}_0$ is the $m \times m$ scale parameter matrix and ν is the scalar shape, or degree of belief, parameter. The mean of **T** is $\mathbf{T}_0^{-1}/(\nu - m - 1)$ (Johnson and Kotz, 1972).

Let $\mathbf{G}_0$ (ν_g), $\mathbf{D}_{0_i}$ (ν_{d_i}), and $\mathbf{R}_{0_j}$ (ν_{r_j}) be the scale (shape) parameters for genetic, group i of uncorrelated random, and group j of residual (co)variances, respectively. In addition, let $\nu_g^* = \nu_g - d_0 - 1$, $\nu_{d_i}^* = \nu_{d_i} - d_i - 1$, and $\nu_{r_i}^* = \nu_{r_i} - t_i - 1$. In the MTGSAM programs the scale matrix is calculated such that the expected value of the prior distribution of the (co)variance matrix is equal to the value entered by the user. For example, for the genetic (co)variance matrix, if the mean value specified by the user is $\mathbf{G}^*$, then $\mathbf{G}_0^{-1} = \nu_g^* \mathbf{G}^*$.

Joint Posterior Density

Let $\mathbf{u}_i' = \begin{bmatrix} \mathbf{u}_{i_1}' & \mathbf{u}_{i_2}' & \dots & \mathbf{u}_{i_{d_i}}' \end{bmatrix}$ for i = 0, ..., γ, then $\mathbf{S}_0 = \left\{ s_{0_{j,k}} \right\}$, where $s_{0_{j,k}} = \mathbf{u}_{0_j}' \mathbf{A}^{-1} \mathbf{u}_{0_k}$. Similarly, $\mathbf{S}_i = \left\{ s_{i_{j,k}} \right\}$, where $s_{i_{j,k}} = \mathbf{u}_{i_j}' \mathbf{u}_{i_k}$, for i = 1, 2, ..., γ. Further, define $\mathbf{e}_{i,j}$ as the vector of residual effects for the traits in residual group i and record j, then $\mathbf{Q}_i = \sum_{j=1}^{q_i} \mathbf{e}_{i,j}\mathbf{e}_{i,j}'$ where q_i is the number of records represented in residual group i.

The joint posterior density, i.e., the density of the parameters given the data and the prior information, can be written as the product of the prior distributions and the likelihood function. Then the joint posterior density is

$$f\left(\beta,\mathbf{u},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,...,\mathbf{D}_\gamma,\mathbf{R} \mid \mathbf{y},\nu_g,\mathbf{G}_0,\nu_{d_1},\mathbf{D}_{0_1},...,\nu_{d_\gamma},\mathbf{D}_{0_\gamma},\nu_{r_1},\mathbf{R}_{0_1},...,\nu_{r_\rho},\mathbf{R}_{0_\rho}\right)$$

$$\propto\; f(\mathbf{y} \mid \beta,\mathbf{u},\mathbf{R}) \times f\left(\mathbf{u} \mid \mathbf{G},\mathbf{D}_1,...,\mathbf{D}_\gamma\right) \times f(\beta) \times f\left(\mathbf{G} \mid \nu_g,\mathbf{G}_0\right) \times \prod_{i=1}^{\gamma}\left[f\left(\mathbf{D}_i \mid \nu_{d_i},\mathbf{D}_{0_i}\right)\right] \times \prod_{i=1}^{\rho}\left[f\left(\mathbf{R}_i \mid \nu_{r_i},\mathbf{R}_{0_i}\right)\right]$$

$$\propto\; f(\mathbf{y} \mid \beta,\mathbf{u},\mathbf{R}) \times f\left(\mathbf{u}_0 \mid \mathbf{G}\right) \times \prod_{i=1}^{\gamma}f\left(\mathbf{u}_i \mid \mathbf{D}_i\right) \times f(\beta) \times f\left(\mathbf{G} \mid \nu_g,\mathbf{G}_0\right) \times \prod_{i=1}^{\gamma}\left[f\left(\mathbf{D}_i \mid \nu_{d_i},\mathbf{D}_{0_i}\right)\right] \times \prod_{i=1}^{\rho}\left[f\left(\mathbf{R}_i \mid \nu_{r_i},\mathbf{R}_{0_i}\right)\right].$$

$$f\left(\beta, \mathbf{u}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_\gamma, \mathbf{R} \mid \mathbf{y}, \nu_g, \mathbf{G}_0, \nu_{d_1}, \mathbf{D}_{0_1}, ..., \nu_{d_\gamma}, \mathbf{D}_{0_\gamma}, \nu_{r_1}, \mathbf{R}_{0_1}, ..., \nu_{r_\rho}, \mathbf{R}_{0_\rho}\right)$$

$$\propto |\mathbf{R}|^{-1/2} \times \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta - \mathbf{Zu})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\beta - \mathbf{Zu})\right\}$$

$$\times \; |\mathbf{G}|^{-n/2} \times \exp\left\{-\frac{1}{2}\mathbf{u}_0'(\mathbf{G}^{-1}\otimes\mathbf{A}^{-1})\mathbf{u}_0\right\} \times \prod_{i=1}^{\gamma}\left[|\mathbf{D}_i|^{-n_i/2} \times \exp\left\{-\frac{1}{2}\mathbf{u}_i'\left(\mathbf{D}_i^{-1}\otimes\mathbf{I}_{n_i}\right)\mathbf{u}_i\right\}\right]$$

$$\times \; |\mathbf{G}|^{-\frac{1}{2}(\nu_g+d_0+1)} \times \exp\left\{\mathrm{tr}\left(-\frac{\nu_g^*}{2}\mathbf{G}_0\mathbf{G}^{-1}\right)\right\} \times \prod_{i=1}^{\gamma}\left[|\mathbf{D}_i|^{-\frac{1}{2}(\nu_{d_i}+d_i+1)} \times \exp\left\{\mathrm{tr}\left(-\frac{\nu_{d_i}^*}{2}\mathbf{D}_{0_i}\mathbf{D}_i^{-1}\right)\right\}\right]$$

$$\times \; \prod_{i=1}^{\rho}\left[|\mathbf{R}_i|^{-\frac{1}{2}(\nu_{r_i}+t_i+1)} \times \exp\left\{\mathrm{tr}\left(-\frac{\nu_{r_i}^*}{2}\mathbf{R}_{0_i}\mathbf{R}_i^{-1}\right)\right\}\right]$$

$$\propto |\mathbf{G}|^{-\frac{1}{2}(n+\nu_g+d_0+1)} \times \exp\left\{-\frac{1}{2}\mathbf{u}_0'(\mathbf{G}^{-1}\otimes\mathbf{A}^{-1})\mathbf{u}_0 + \mathrm{tr}\left(\nu_g^*\mathbf{G}_0\mathbf{G}^{-1}\right)\right\}$$

$$\times \; \prod_{i=1}^{\gamma}\left[|\mathbf{D}_i|^{-\frac{1}{2}(n_i+\nu_{d_i}+d_i+1)} \times \exp\left\{-\frac{1}{2}\left(\mathbf{u}_i'\left(\mathbf{D}_i^{-1}\otimes\mathbf{I}_{n_i}\right)\mathbf{u}_i + \mathrm{tr}\left(\nu_{d_i}^*\mathbf{D}_{0_i}\mathbf{D}_i^{-1}\right)\right)\right\}\right]$$

$$\times \; \prod_{i=1}^{\rho}\left[|\mathbf{R}_i|^{-\frac{1}{2}(q_i+\nu_{r_i}+t_i+1)} \times \exp\left\{\mathrm{tr}\left(-\frac{\nu_{r_i}^*}{2}\mathbf{R}_{0_i}\mathbf{R}_i^{-1}\right)\right\}\right]$$

$$\times \; \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta - \mathbf{Zu})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\beta - \mathbf{Zu})\right\}. \qquad [1]$$

An alternative form of the density is:

$$f\left(\beta, \mathbf{u}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_\gamma, \mathbf{R} \mid \mathbf{y}, \nu_g, \mathbf{G}_0, \nu_{d_1}, \mathbf{D}_{0_1}, ..., \nu_{d_\gamma}, \mathbf{D}_{0_\gamma}, \nu_{r_1}, \mathbf{R}_{0_1}, ..., \nu_{r_\rho}, \mathbf{R}_{0_\rho}\right)$$

$$\propto |\mathbf{G}|^{-\frac{1}{2}(n+\nu_g+d_0+1)} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left(\nu_g^*\mathbf{G}_0+\mathbf{S}_0\right)\mathbf{G}^{-1}\right)\right\}$$

$$\times \; \prod_{i=1}^{\gamma}\left[|\mathbf{D}_i|^{-\frac{1}{2}(n_i+\nu_{d_i}+d_i+1)} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left(\nu_{d_i}^*\mathbf{D}_{0_i}+\mathbf{S}_i\right)\mathbf{D}_i^{-1}\right)\right\}\right]$$

$$\times \; \prod_{i=1}^{\rho}\left[|\mathbf{R}_i|^{-\frac{1}{2}(q_i+\nu_{r_i}+t_i+1)} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left(\nu_{r_i}^*\mathbf{R}_{0_i}+\mathbf{Q}_i\right)\mathbf{R}_i^{-1}\right)\right\}\right]. \qquad [2]$$

Fully Conditional Densities

A set of fully conditional densities is required to implement GS. These densities are required for each scalar element or subvector of elements in the vector of parameters to be estimated. Each fully conditional density corresponds to the distribution of the specific parameter(s) conditional on all other parameters in the model and the data; i.e., it is the distribution of the specific parameter if the values of the remaining parameters were known. The fully conditional densities can be derived from the joint posterior density, i.e., from [1] or [2], by ignoring all terms not involving the parameter(s) of interest, then treating the parameters considered to be known as constants and reorganizing terms retained into the density kernel of a known distribution for the parameter of interest. The kernel of a distribution is the part of the function containing the random variable, which remains when all constants are discarded.

Fixed and Random Effects. First, the conditional distribution of the fixed and random effects, $\mathbf{s} = [\beta' \; \mathbf{u}']'$, will be obtained. A useful result from the form of the MME is

$$\mathbf{s'Cs} = [\beta' \; \mathbf{u}']\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z} + \Sigma^{-1} \end{bmatrix}\begin{bmatrix} \beta \\ \mathbf{u} \end{bmatrix}$$

$$= \mathbf{s'W'R^{-1}Ws} + \mathbf{u'}\Sigma^{-1}\mathbf{u}, \text{ where}$$

$$\mathbf{W} = [\mathbf{X} \; \mathbf{Z}].$$

Then, the conditional distribution of the fixed and random effects, **s**, can be written as (Gianola and Fernando, 1986; Gianola et al., 1990):

$$f\ (\mathbf{s}\,|\,\mathbf{G},\mathbf{D_1},\mathbf{D_2},...,\mathbf{D_\gamma},\mathbf{R},\mathbf{y})$$

$$\propto \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{Ws})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Ws})\right\}\ \times\ \exp\left\{-\frac{1}{2}\mathbf{u}'\Sigma^{-1}\mathbf{u}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}((\mathbf{s}-\tilde{\mathbf{s}})'\mathbf{C}(\mathbf{s}-\tilde{\mathbf{s}}))\right\}\ ,$$

where

$$\tilde{\mathbf{s}}\ =\ \mathbf{C}^{-1}\mathbf{W}'\mathbf{R}^{-1}\mathbf{y}.$$

This is the kernel of a normal density, and therefore

$$\mathbf{s}\,|\,\mathbf{G},\mathbf{D_1}\mathbf{D_2},...,\mathbf{D_\gamma},\mathbf{R},\mathbf{y}\ \sim\ \mathrm{N}\ (\tilde{\mathbf{s}},\ \mathbf{C}^{-1})\ . \qquad [3]$$

Using results for the conditional multivariate normal distribution, the fully conditional distribution for a scalar or subvector element $\mathbf{s_1}$ of $\mathbf{s}$, where $\mathbf{s}' = \begin{bmatrix}\mathbf{s}_1' & \mathbf{s}_2'\end{bmatrix}$, is

$$\mathbf{s_1}\,|\,\mathbf{s_2},\mathbf{G},\mathbf{D_1},\mathbf{D_2},...,\mathbf{D_\gamma},\mathbf{R},\mathbf{y}\ \sim\ \mathrm{N}(\mathbf{C}_{11}^{-1}(\mathbf{r_1}-\mathbf{C}_{12}\mathbf{s_2}),\mathbf{C}_{11}^{-1}),\ \ [4]$$

and

$$\mathbf{C}\ =\ \begin{bmatrix}\mathbf{C}_{11} & \mathbf{C}_{12}\\ \mathbf{C}_{21} & \mathbf{C}_{22}\end{bmatrix}\ .$$

Blocked Random Effects. The MTGSAM programs use a blocked GS algorithm to generate all genetic effects for an animal as well as associated uncorrelated random effects simultaneously. When the posterior correlations of effects are large, the means of fully conditional distributions for the RV involved may be highly dependent, resulting in a limited practical sampling range for that random variable given the current value of a correlated variable. For example, for dairy cattle data with multiple lactations recorded, often a permanent environmental effect (**PE**) is included to account for non-genetic, animal-specific, random effects. The PE levels would correspond to the animal levels, that is, the level for animal identification and PE effect would be coded in the same field of the original data set. If the genetic and PE effects are generated individually, then the range of one effect, say PE, is limited by the current value of the genetic effect. This in turn leads to the animal effect being limited by the current value of the PE. As a result, the consecutive samples of the same parameter may be highly correlated, which reduces the efficiency and rate of convergence of the GS algorithm. This phenomenon is known as slow mixing, where the

mixing rate refers to the rate that the Gibbs sampler moves through all of the potential values in the distribution.

The MTGSAM programs for that case generate the PE effects simultaneously with the genetic effects. In general, the programs block and generate simultaneously any uncorrelated random effect that is coded in the same column as an animal or correlated genetic effect with the genetic effects. Generating correlated variables in blocks should increase the mixing rate of the Gibbs sampler and reduce the correlations among samples drawn. This method reduces correlations among the samples drawn, but, because not all correlated variables are sampled simultaneously, blocking may not completely eliminate the correlations within a Gibbs chain (Liu et al., 1994; Van Tassell et al., 1994).

A correlation among variables that are assumed to be statistically uncorrelated can be induced through the data structure, i.e., through the least squares part of the equations. If we reconsider the dairy example, genetic and PE effects are assumed to be uncorrelated, leading to a block diagonal structure in the covariance matrix, but for animals with records observed there are non-zero elements that bridge the block-diagonal structure; as a result there is a posterior correlation among the genetic and PE effects for an animal. In models with maternal genetic and PE effects, those effects are blocked together, i.e., the maternal PE effect is blocked with the maternal genetic effect because those effects are correlated due to the data structure. Previous results for a maternal effects model applied to Simmental weaning weight data support this concept, although the difference between blocked and scalar algorithms was relatively small (Van Tassell et al., 1994).

To derive the form of the fully conditional distribution for the block of effects, additional definitions are needed. Let **P** be a permutation matrix such that

$$\mathbf{Ps}\ =\ \begin{bmatrix}\mathbf{P_i s}\\ \mathbf{P_{-i} s}\end{bmatrix}\ =\ \begin{bmatrix}\mathbf{s_i}\\ \mathbf{s_{-i}}\end{bmatrix}\ ,$$

where $\mathbf{s_i}$ are the random effects in the block, and $\mathbf{s_{-i}}$ are the remaining fixed and random effects. The matrix **P** can be thought of as an identity matrix with the order of the columns altered so that **Ps** contains the same elements as **s** but in a different order. Then application of a property of the conditional normal distribution to the fully conditional distribution of the fixed and random effects in [3] results in the fully conditional distribution of the block of random effects:

$$\mathbf{s_i}\,|\,\mathbf{s_{-i}}\mathbf{G},\mathbf{D_1},\mathbf{D_2},...,\mathbf{D_\gamma},\mathbf{R},\mathbf{y}\ \sim\ \mathrm{N}(\tilde{\mathbf{s}}_\mathbf{i},(\mathbf{P_i C P}_\mathbf{i}')^{-1})\ ,\ \ [5]$$

where

$$\tilde{\mathbf{s}}_\mathbf{i}\ =\ (\mathbf{P_i C P}_\mathbf{i}')^{-1}(\mathbf{P_i r}-\mathbf{P_i C P}_{-\mathbf{i}}'\mathbf{s_{-i}})\ .$$

Although this form appears complex, it is simply a precise representation of [4], where the permutation of elements is shown here explicitly. The (co)variance matrix is composed of the appropriate elements of the coefficient matrix corresponding to the rows and columns for the elements in the block, and the mean is a function of that matrix and the right-hand sides for the blocked effects adjusted for the off-diagonal elements of the effects not generated in that block. The adjustments to the right-hand sides are based on the rows of the coefficient matrix for those blocked elements with the columns removed for the elements included in that block. The remaining matrix is multiplied by the vector of current values for the fixed and random effects not included in the block and subtracted from the appropriate element.

Missing Data. To allow for missing traits, the residual and uncorrelated random effects for missing traits must be generated to calculate quadratic forms for these effects and for generating residual (co)variance matrices. There will only be missing uncorrelated random levels when the same effect is used in several traits, e.g., litter or PE. The missing value is calculated using the same fully conditional distribution as the effects that have observations, i.e.,

from [5]. The missing residuals are generated using the conditional normal distribution, i.e., the residuals are calculated for the observed traits and the missing residuals are generated using GS based on the current values of the residual (co)variances for that group. To specify the form of the fully conditional distribution of the missing residual effects, assume, without loss of generality, the vector of residuals can be partitioned as

$$\mathbf{e}_{i,j} = \begin{bmatrix} \mathbf{e}_m \\ \mathbf{e}_o \end{bmatrix} ,$$

where $\mathbf{e}_m$ is the subvector of missing residuals and $\mathbf{e}_o$ is the subvector of residuals for observed traits. For group i of residual effects, the residuals are distributed normally:

$$\mathbf{e}_i \,|\, \mathbf{R}_i \;\sim\; N(\mathbf{0}, \mathbf{R}_i) \ .$$

Let

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{mm} & \mathbf{R}_{mo} \\ \mathbf{R}_{om} & \mathbf{R}_{oo} \end{bmatrix} ;$$

then, using the form of the conditional normal distribution,

$$\mathbf{e}_m \,|\, \mathbf{e_o}, \mathbf{s}, \mathbf{G}, \mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_\gamma, \mathbf{R}, \mathbf{y} \;\sim\; N\!\left(\mathbf{R}_{mo}\mathbf{R}_{oo}^{-1}\mathbf{e}_o, \mathbf{R}_{mm} - \mathbf{R}_{mo}\mathbf{R}_{oo}^{-1}\mathbf{R}_{om}\right) \ . \tag{6}$$

(Co)variance Matrices. The fully conditional distributions of the (co)variance matrices can be derived using form [2] of the joint posterior distribution. The fully conditional distribution of the genetic (co)variance matrix is

$$f\left(\mathbf{G} \,|\, \mathbf{s}, \mathbf{G}_0 \nu_g, \mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_\gamma, \mathbf{R}, \mathbf{y}\right) \;\propto\; |\mathbf{G}|^{-\frac{1}{2}(n + \nu_g + d_0 + 1)} \times \exp\!\left\{-\frac{1}{2}\mathrm{tr}\!\left((\nu_g^* \mathbf{G}_0 + \mathbf{S}_0)\mathbf{G}^{-1}\right)\right\} \ .$$

This is the kernel of an IW density, so that

$$\mathbf{G} \,|\, \mathbf{s}, \mathbf{G}_0 \nu_g, \mathbf{D}_1 \mathbf{D}_2, ..., \mathbf{D}_\gamma, \mathbf{R}, \mathbf{y} \;\sim\; IW\!\left((\nu_g^* \mathbf{G}_0 + \mathbf{S}_0)^{-1}, n + \nu_g\right) \ . \tag{7}$$

The fully conditional distribution for the matrix of (co)variances for each group of uncorrelated random effects is

$$f\left(\mathbf{D}_i \,|\, \mathbf{s}, \mathbf{d_{0_i}}, \nu_{\mathbf{d_i}}, \mathbf{G}, \mathbf{D}_1 \mathbf{D}_2, ..., \mathbf{D}_{i-1}, \mathbf{D}_{i+1}, ..., \mathbf{D}_\gamma, \mathbf{R}, \mathbf{y}\right) \propto |\mathbf{D}_i|^{-\frac{1}{2}(n_i + \nu_{d_i} + d_i + 1)} \times \exp\!\left\{-\frac{1}{2}\mathrm{tr}\!\left((\nu_{d_i}^* \mathbf{D}_{0_i} + \mathbf{S}_i)\mathbf{D}_i^{-1}\right)\right\} \ .$$

This is also the kernel of an IW density, so that

$$\mathbf{D}_i \,|\, \mathbf{s}, \mathbf{D}_{0_i}, \nu_{d_i}, \mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_{i-1}, \mathbf{D}_{i+1}, ..., \mathbf{D}_\gamma, \mathbf{R}, \mathbf{y} \;\sim\; IW\!\left((\nu_{d_i}^* \mathbf{D}_{0_i} + \mathbf{S}_i)^{-1}, n_i + \nu_{d_i}\right) \ . \tag{8}$$

The fully conditional distribution for the matrix of (co)variances for each group of residual effects is

$$f\left(\mathbf{R}_i \,|\, \mathbf{e}_i, \mathbf{R}_{0_i}, \nu_{r_i}, \mathbf{s}, \mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_\gamma, \mathbf{R}_1, ..., \mathbf{R}_{i-1}, \mathbf{R}_{i+1}, ..., \mathbf{R}_\rho, \mathbf{y}\right) \;\propto\; |\mathbf{R}_i|^{-\frac{1}{2}(q_i + \nu_{r_i} + t_i + 1)} \times \exp\!\left\{-\frac{1}{2}\mathrm{tr}\!\left((\nu_{r_i}^* \mathbf{R}_{0_i} + \mathbf{Q}_i)\mathbf{R}_i^{-1}\right)\right\} \ .$$

Again, this function is the kernel of an IW density so that

$$\mathbf{R}_i \mid \mathbf{e}_i, \mathbf{R}_{0_i}, \nu_{r_i}, \mathbf{s}, \mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_\gamma, \mathbf{R}_1, ..., \mathbf{R}_{i-1}, \mathbf{R}_{i+1}, ..., \mathbf{R}_\rho, \mathbf{y} \sim IW\left(\left(\nu_{r_i}^* \mathbf{R}_{0_i} + \mathbf{Q}_i\right)^{-1}, q_i + \nu_{r_i}\right) . \qquad [9]$$

Implementation of the Gibbs Sampler

Using the fully conditional distributions, i.e., [4] to [9], the GS algorithm used can be outlined as follows:

1. Input or calculate starting values for all variables.
   a. Starting (co)variances supplied by the user are used for those components. The starting values correspond to the means of prior distributions if using informative prior distributions for the (co)variances.
   b. Gauss-Seidel iteration using the starting VC is used to calculate starting values for fixed and random effects. The user specifies maximum number of rounds of iteration and convergence criterion.
2. Generate fixed effects from [4].
3. Generate genetic effects and blocked uncorrelated random effects from [5].
4. Generate uncorrelated random effects not in a block from [4].
5. Calculate residual effects for traits with observations, and generate missing residuals from [6].
6. Calculate quadratics for genetic effects, $\mathbf{S}_0$, from $\mathbf{u}_{0_i}' \mathbf{A}^{-1} \mathbf{u}_{0_j}$.
7. Generate $\mathbf{G}$ from [7].
8. Calculate quadratics for each group of uncorrelated random effects, $\mathbf{S}_i$, from $\mathbf{u}_{i_j}' \mathbf{u}_{i_k}$.
9. Generate each $\mathbf{D}_i$ from [8].
10. Calculate quadratics for each group of residual effects, $\mathbf{Q}_i$, from $\mathbf{e}_{i,j} \mathbf{e}_{i,j}'$.
11. Generate each $\mathbf{R}_i$ from [9].
12. Repeat steps 2 through 11 many times.

Estimation of Posterior Means

The mean of a parameter is a useful and convenient point estimate. There are two basic methods used by MTGSAM to estimate the posterior means, depending on the variable. The first is based on the average of the conditional expected values of the parameter, and the second is based on the average of the sampled values. For variables for which it is possible, the mean of the conditional expectation of the parameter is used. The conditional expected value used in the calculation corresponds to the expected value of the fully conditional distribution for that parameter from which the new value was sampled, i.e., conditional on the values of the other parameters when that sample was drawn. This form of the estimate is used when possible because it is the Rao-Blackwell estimator; i.e., it is the minimum variance estimator. There are cases in which the conditional expectation of the parameter is not known, specifically, for functions of parameters. These include phenotypic variance, correlations,

heritabilities (or other fractions of phenotypic variance), and linear combinations of fixed and random effects (i.e., estimable or predictable functions). The mean of the sampled values is used in that case. For example, to calculate the mean of the difference between two fixed effects, the observed difference would be calculated in each round and that value would be averaged.

Variance Components. The MTGSAM posterior mean estimate for (co)variance components is based on the expected value of the IW RV. Recall that if $\mathbf{T} \sim IW(\mathbf{V}, \nu)$, then $E(\mathbf{T}) = \mathbf{V}^{-1}/(\nu - m - 1)$. Therefore, from [7], [8], and [9] the expected values for (co)variance matrices for a given round of Gibbs sampling are calculated as

$$E\left(\mathbf{G} \mid \mathbf{S}_0, \mathbf{G}_0, \nu_g\right) = \frac{\nu_g^* \mathbf{G}_0 + \mathbf{S}_0}{\left(n + \nu_g - d_0 - 1\right)} ,$$

$$E\left(\mathbf{D}_i \mid \mathbf{S}_i, \mathbf{D}_{0_i}, \nu_{d_i}\right) = \frac{\nu_{d_i}^* \mathbf{D}_{0_i} + \mathbf{S}_i}{\left(n_i + \nu_{d_i} - d_i - 1\right)},$$

and

$$E\left(\mathbf{R}_i \mid \mathbf{Q}_i, \mathbf{R}_{0_i}, \nu_{r_i}\right) = \frac{\nu_{r_i}^* \mathbf{R}_{0_i} + \mathbf{Q}_i}{\left(q_i + \nu_{r_i} - t_i - 1\right)}.$$

The mean of a (co)variance component is calculated as the average of these expected values over the length of the post burn-in chain.

Functions of Variance Components. Functions of (co)variance components considered by MTGSAM include the phenotypic (co)variances (sum of appropriate genetic, uncorrelated random and residual (co)variances), correlations, and fraction of phenotypic variance accounted for by a particular VC (e.g., heritability, fraction due to uncorrelated random effects, or fraction due to residual effects). Because the conditional distributions of these functions cannot be written in closed form, MTGSAM calculates the estimates of the posterior means for these parameters as the means of the functions calculated using the sampled (co)variance components in each post burn-in round of GS.

Fixed and Random Effects. Recall that the fully conditional distribution of a fixed or random effect is a normal distribution. The posterior means for those parameters are estimated in MTGSAM as the averages of the means of the normal distribution that the parameters are sampled from in each post burn-in round of GS.

Functions of Fixed and Random Effects. The MTGSAM programs consider two different linear combinations of fixed and random effects: those based on a single effect and those based on multiple fixed or random effects. A single element may be specified if a

particular effect is of interest. An example of the linear function might be a case in which the distribution of genetic effects for a specified group of animals is wanted, but the amount of information generated would be prohibitive for all animals (i.e., using the option in the program to write sample information for fixed and random effects is impractical). When the function contains only one effect, the program uses the parametric estimate of the mean of that function, i.e., the average of the means of the fully conditional normal distributions. When the contrast includes multiple effects, a nonparametric approach must be used to estimate the mean. The mean is estimated as the average of the sampled values, which is determined by calculating the linear combination of sampled fixed and random effects specified by the function after burn-in.

Gauss-Seidel Iteration

The MTGSAM programs use Gauss-Seidel iteration (**GSI**) to obtain starting values for fixed and random effects for GS based on the starting (co)variance components. In addition, the programs can be used to obtain solutions to the MME using GSI without GS. Because of unique memory usage strategies, the MTGSAM programs may be able to solve larger systems of equations than many other general programs. The least squares part of the equations is half-stored using sparse matrix storage. The inverse of the genetic (co)variance matrix (i.e., $\Sigma_0^{-1} = \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}$) is not explicitly added to the least squares portion of the equations; instead, one copy of $\mathbf{A}^{-1}$ is full sparse stored. The relative memory requirement for explicitly adding the half-stored Σ_0^{-1} matrix to the least squares coefficients is $n^2/2$ to 1, where n is the number of genetic traits (i.e., number of traits and correlated genetic effects). Therefore, if $n \geq 2$, then this storage scheme will be more memory efficient than half-storing the augmented equations. The algorithm used for GSI is very similar to that used for GS; the update to a variable or vector of variables in GSI is the same as the algorithm to calculate the mean of the normal distributions for fixed and random effects in GS.

Data Simulation

Data were simulated using a bivariate form of the scheme described by Sorensen and Kennedy (1984) and applied by Van der Werf and De Boer (1990). Fifty replicates were generated; each contained 400 animals with 50 males and 50 females born in each of four generations. Ten males were chosen in each generation as sires of the next generation based on the phenotypic value for trait one; all females were chosen to be parents. Each male was mated to five females. Each mating generated a male and a female progeny. The simulation model included additive genetic and

residual effects for each trait. The genetic variance in the base generation was 1.0 for both traits; the genetic correlation was .3. Residual variance was 1.0 for both traits; the residual correlation was .1.

Data Analysis

The model used for data analysis included a fixed mean for each trait and random genetic and residual effects. The relationships among all animals were used to calculate the numerator relationship matrix. Point estimates were obtained for the genetic and residual variance for each trait and for the genetic and residual covariances.

The MTGSAM programs were used to generate files of samples and posterior estimates of variance components. Two different sets of prior distributions were used for the (co)variance components: flat prior distributions and IW distributions with shape parameters of 10 and a mean equal to the values used in the simulations. The Gibbs sampler was run 15,000 rounds with the first 2,000 rounds discarded for burn-in. The posterior mean was estimated for (co)variance components, heritabilities, and correlations.

For comparison, the MTDFREML programs (Boldman et al., 1995) were also used to estimate variance components. Estimates were obtained for (co)variance components, heritabilities, and correlations. The analyses were restarted without the simplex information at least three times; if the log-likelihood of the best value changed by at least .01 between the last two runs, then restarts were continued until the change was less than .01.

Burn-in and Thinning Analyses

The burn-in period and thinning interval were evaluated for two randomly chosen data sets using both informative and flat prior distributions; i.e., four Gibbs chains were evaluated. The analyses were based on GS output for chains run 15,000 rounds, with the first 2,000 rounds samples discarded as burn-in period. The Gibbsit program of Raftery and Lewis (1994), based on the algorithm described by Raftery and Lewis (1992), was used to evaluate burn-in for all (co)variance components and functions of VC. The Gibbsit program was also used to calculate thinning intervals to determine the frequency of retaining sampled values so that those samples are uncorrelated for each parameter. Lag correlations among samples were also calculated to further evaluate thinning intervals.

Monte Carlo error was estimated by calculating the variance of the parameters sampled 50-rounds apart and dividing this variance by 260, where 260 is the number of thinned samples. The square root of this value is an approximation to the standard deviation of the error associated with the length of the Gibbs chain.

Table 1. Means and empirical standard deviations of estimates and correlations between estimates for 50 simulated populations using Gibbs sampling with informative priors (GS-I) and flat priors (GS-F) and REML

Parameter[a]	Means			Standard deviations			Correlations[b]		
	GS-I	GS-F	REML	GS-I	GS-F	REML	GS-I GS-F	GS-I REML	GS-F REML
$\sigma^2_{G_1}$	1.029	1.028	.997	.178	.198	.194	.984	.985	.999
σ_{G_1,G_2}	.302	.295	.288	.133	.151	.147	.983	.986	.999
$\sigma^2_{G_2}$	1.052	1.030	.998	.207	.229	.224	.944	.949	.996
h^2_1	.513	.503	.500	.066	.072	.073	.971	.974	.998
$r_{G_1G_2}$	.293	.297	.295	.124	.161	.154	.980	.990	.995
h^2_2	.515	.496	.494	.079	.090	.089	.938	.944	.994
$\sigma^2_{R_1}$	.963	.998	.988	.134	.139	.137	.970	.971	.999
σ_{R_1,R_2}	.095	.104	.103	.099	.105	.104	.985	.988	.999
$\sigma^2_{R_2}$	.972	1.022	1.011	.169	.181	.179	.953	.955	.997
r_{R_1,R_2}	.094	.101	.103	.106	.107	.106	.981	.984	.999
$\sigma^2_{P_1}$	1.997	2.031	1.984	.160	.171	.166	.998	.998	.999
σ_{P_1,P_2}	.398	.401	.392	.140	.150	.147	.997	.998	.999
$\sigma^2_{P_2}$	2.030	2.057	2.010	.176	.187	.182	.991	.993	.999
r_{P_1,P_2}	.197	.195	.195	.064	.068	.068	.997	.998	.999

[a]G_1, G_2, R_1, R_2, P_1, P_2 are genetic effects for traits 1 and 2, residual effects for traits 1 and 2, and phenotypic measurements for traits 1 and 2, respectively; σ^2, σ, and r represent variance, covariance, and correlation; and h^2_1 and h^2_2 are heritabilities for traits 1 and 2.

[b]Correlations between estimates for the two methods listed in the column heading.

Results and Discussion

The Gibbsit program of Raftery and Lewis (1994) recommended burn-in periods of 30 to 56 rounds. A conservative burn-in period of 2,000 rounds was used. Gibbsit calculated thinning intervals of 12 to 15 rounds. Results for the correlations among samples indicated that the thinning intervals recommended by Gibbsit were not large enough. Correlations for samples drawn 50 rounds apart ranged from −.042 to .093, with an average of .033. The correlations of most parameters were less than .05 at a 50-round sampling interval.

The maximums of square roots of the Monte Carlo errors across the four chains evaluated were .015, .018, .012, .009, .012, and .008 for the genetic variances of traits one and two, genetic covariance, residual variances for traits one and two, and residual covariance, respectively. For heritabilities for traits one and two, genetic correlation, and residual correlation, the maximums of the square roots of the Monte Carlo errors were .005, .006, .009, and .007. These small Monte Carlo errors provide strong evidence that the 13,000 rounds of post burn-in Gibbs sampling used in this study were sufficient to obtain precise estimates of posterior means.

Means and empirical standard deviations of posterior mean estimates for GS using informative prior distributions for VC (**GS-I**), GS using flat prior distributions (**GS-F**), and REML are presented in Table 1. Although theoretically all three methods are biased, there is little evidence of systematic bias based on the mean point estimates for the 50 replicates. The means in some cases were statistically different from the simulated values, but there were no systematic differences. Significance was mainly due to the relatively small variation in point estimates even when based on only 50 populations with 400 animals. It should be noted that, if prior distributions were used with means different from the true values, then the bias and mean squared errors of the estimates of the posterior means would likely increase. Bias and mean squared errors for posterior mean estimates increased as the difference between the mean of the prior distribution and the true value increased using a similar univariate simulation (Van Tassell et al., 1995).

Correlations among the parameter estimates are also presented in Table 1. Although large, the correlations between GS-I and GS-F and between GS-I and REML were less than the correlations between GS-F and REML. The biggest difference between the GS-I point estimates and the GS-F or REML estimates was the smaller variance of the estimates due to the informative priors. This is not surprising, because this estimate combines information from the prior distribu-
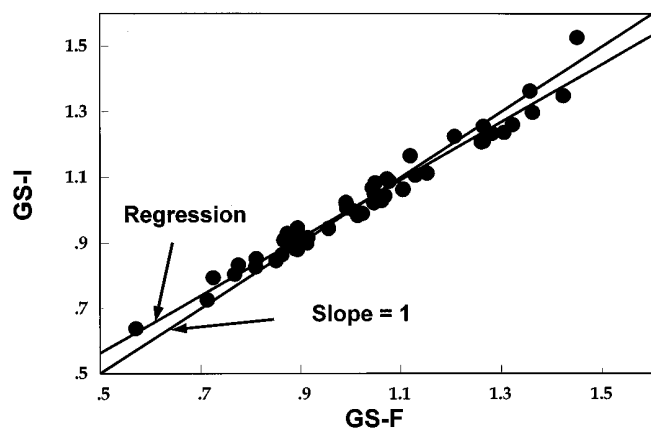
Figure 1. Comparison of mean posterior estimates for genetic variance of trait 1 using Gibbs sampling with informative (GS-I) and flat (GS-F) prior distributions for variance components. Regression of GS-I on GS-F estimates is presented with a line with a slope of 1 for comparison.
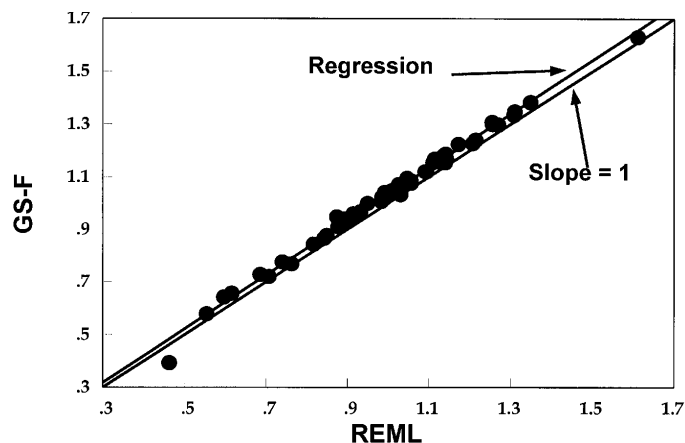

Figure 2. Comparison of mean posterior estimates for genetic variance of trait 2 using Gibbs sampling with flat prior distributions for variance components (GS-F) and REML estimates of genetic variance for trait 2. Regression of GS-F on REML estimates is presented with a line with a slope of 1 for comparison.

tion and from the data; the means of these two pieces of information, in this case, should have similar means. However, with the use of poor prior distributions, i.e., with means very different from the population, the variance of the estimator can increase. As a result of the reduced variance of the GS-I point estimates, the slope of the line regressing GS-I on the other methods was less than one (Figure 1). The effect of the prior distribution is to "regress" the sample information towards the mean specified by the prior distribution. This "regression" is a function of the shape parameter specified for the prior distribution and the number of levels of the random effect (i.e., animals). This result is similar to the expression given by Gianola et al. (1990) for the *maximum a posteriori* (**MAP**) estimate of the variance component:

$$\sigma^2 \;=\; \frac{(n+2)\tilde{\sigma}^2 \;+\; \nu s^2}{n \;+\; \nu \;+\; 2} \;,$$

where $\tilde{\sigma}^2$ is the "data" variance, ν is the shape parameter for prior distribution, and s^2 is the scale parameter. This can be rearranged to

$$\sigma^2 \;=\; \frac{\nu s^2}{n \;+\; \nu \;+\; 2} \;+\; \frac{(n \;+\; 2)}{n \;+\; \nu \;+\; 2}\, \tilde{\sigma}^2 \;=\; a \;+\; b\tilde{\sigma}^2 \;,$$

which in this form makes the regression more obvious. This relationship does not necessarily hold for the mean of the posterior distribution.

Estimates using GS-F and REML were very similar, as the correlations between estimates with the two methods indicate. This similarity agrees well with the form of the MAP estimator because the portion of the estimate from the prior distribution is zero. A plot of GS-F and REML estimates is presented in Figure 2, where the similarity between the estimates can be

seen. The results from the simulated data suggest that the likelihoods were relatively symmetric in these populations, because the posterior mean and REML estimates agreed quite well. However, results from Van Tassell and Pollak (1994) indicate that even with a much larger data set (5,440 records), mean (GS posterior mean) and modal (REML) estimators can be quite different; i.e., that assumption of normality under the large sample properties of likelihood-based inference may not be appropriate.

The MTGSAM programs are intended to be used by those familiar with the concepts of GS, including evaluation of the burn-in period, thinning interval, and convergence. The programs simplify much of the work required to use GS, but users must still make certain that results obtained are sensible.

The MTGSAM programs are available through the Internet or from the authors. The MTGSAM home page is available via the Animal Geneticists Discussion Group (AGDG) home page, which can be accessed from links on the American Society of Animal Science (ASAS) home page.

Implications

The MTGSAM programs implement the GS algorithm for a wide variety of multiple-trait models, allowing Bayesian analysis of (co)variance components without approximations. The programs estimate posterior means (up to Monte Carlo error) and generate sampled values for (co)variances, correlations, heritabilities, fixed and random effects, and linear combinations of fixed and random effects. Posterior distributions can be estimated using the

samples generated by the programs. Gibbs sampling may allow for parameter estimation for problems that were not possible with a likelihood-based analysis. These situations include the use of larger data sets or more complex models.

Literature Cited

Albert, J. H., and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. J. Am. Stat. Assoc. 88:669.

Boldman, K. G., L. A. Kriese, L. D. Van Vleck, C. P. Van Tassell, and S. D. Kachman. 1995. A Manual for Use of MTDFREML. A Set of Programs To Obtain Estimates of Variances and Covariances [Draft]. U.S. Department of Agriculture, Agricultural Research Service.

Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6:721.

Gianola, D., and R. L. Fernando. 1986. Bayesian methods in animal breeding theory. J. Anim. Sci. 63:217.

Gianola, D., and J. L. Foulley. 1990. Variance estimation from integrated likelihoods (VEIL). Genet. Sel. Evol. 22:403.

Gianola, D., S. Im, and F. W. Macedo. 1990. A framework for prediction of breeding value. In: Gianola, D., and K. Hammond (ed.) Advances in Statistical Methods for Genetic Improvement of Livestock. pp 210–238. Springer-Verlag, Berlin, Germany.

Groeneveld, E. 1994. VCE—A multivariate multimodel REML (co)variance component estimation program. Proc. 5th World Congr. Genet. Appl. Livest. Prod. 22:46

Henderson, C. R. 1975. Comparison of alternative sire evaluation methods. J. Anim. Sci. 41:760.

Henderson, C. R. 1984. Application of Linear Models in Animal Breeding. Univ. Guelph, Guelph, ON, Canada.

Hobert, J. P. 1994. Occurrences and consequences of nonpositive Markov chains in Gibbs sampling. Ph.D. Dissertation. Cornell University, Ithaca, NY.

Jensen, J., C. S. Wang, D. A. Sorensen, and D. Gianola. 1994. Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. Acta Agric. Scand. 44:193.

Johnson, N. L., and S. Kotz. 1972. Distributions in Statistics. John Wiley & Sons, New York.

Liu, J. S., W. H. Wong, and A. Kong. 1994. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. Biometrika 81:27.

Meyer, K. 1988. DFREML. A set of programs to estimate variance components under an individual animal model. J. Dairy Sci. 69(Suppl. 2):33.

Meyer, K. 1990. Present status of knowledge about statistical procedures and algorithms to estimate variance and covariance components. Proc. 4th World Cong. Appl. Livest. Prod., Edinburgh, Scotland, XIII:407.

Odell, P. L., and A. H. Feiveson. 1966. A numerical procedure to generate a sample covariance matrix. J. Am. Stat. Assoc. 61:199.

Raftery, A. E., and S. M. Lewis. 1992. How many iterations in the Gibbs sampler? In: Bernardo, J. M., J. O. Berger, A. P. Dawid, and A.F.M. Smith (ed.) Bayesian Statistics, Vol. 4. p 763. Oxford University Press, Oxford, U.K.

Raftery, A. E., and S. M. Lewis. 1994. Gibbsit version 2.0. StatLib file /general/gibbsit. World Wide Web URL is http://lib.stat.cmu.edu/general/gibbsit.

Schaeffer, L. R. 1984. Sire and cow evaluation under multiple trait models. J. Dairy Sci. 67:1567.

Searle, S. R. 1982. Matrix Algebra Useful for Statistics. John Wiley & Sons, New York.

Sorensen, D. A., S. Anderson, D. Gianola, and I. Korsgaard. 1995. Bayesian inference in threshold models using Gibbs sampling. Genet. Sel. Evol. 27:229.

Sorensen, D. A., and B. W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. J. Dairy Sci. 59:1213.

Sorensen, D. A., C. S. Wang, J. Jensen, and D. Gianola. 1994. Bayesian analysis of genetic change due to selection using Gibbs sampling. Genet. Sel. Evol. 26:333.

Tanner, M. 1993. Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions (2nd ed.). Springer-Verlag, New York.

Van der Werf, J.H.J., and I.J.M. de Boer. 1990. Estimation of additive genetic variance when base populations are selected. J. Anim. Sci. 68:3124.

Van Tassell, C. P. 1994. The use of Gibbs sampling for variance component estimation with simulated and weaning weight data using animal and maternal effects models. Ph.D. Thesis. Cornell University, Ithaca, NY.

Van Tassell, C. P., G. Casella, and E. J. Pollak. 1994. Comparison of efficiency for blocked and scalar Gibbs sampling algorithms. J. Anim. Sci. 72(Suppl. 1):58(Abstr.).

Van Tassell, C. P., G. Casella, and E. J. Pollak. 1995. Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. J. Dairy Sci. 78:678.

Van Tassell, C. P., and E. J. Pollak. 1994. Estimates of (co)variance components for weaning weight used in national beef cattle genetic evaluations. J. Anim. Sci. 72(Suppl. 1):148(Abstr.).

Van Tassell, C. P. and L. D. Van Vleck. 1995. A Manual for Use of MTGSAM. A Set of FORTRAN Programs To Apply Gibbs Sampling to Animal Models for Variance Component Estimation [Draft]. U.S. Department of Agriculture, Agricultural Research Service.

Wang, C. S., D. Gianola, D. A. Sorensen, J. Jensen, A. Christensen, and J. J. Rutledge. 1994a. Response to selection for litter size in Danish Landrace pigs; a Bayesian analysis. Theor. Appl. Genet. 88:220.

Wang, C. S., J. J. Rutledge, and D. Gianola. 1993. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. Genet. Sel. Evol. 25:41.

Wang, C. S., J. J. Rutledge, and D. Gianola. 1994b. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. Genet. Sel. Evol. 26:91.

Citations

This article has been cited by 14 HighWire-hosted articles:
http://jas.fass.org#otherarticles